Treewidth-based conditions for exactness of the Sherali-Adams and Lasserre relaxations

Martin J. Wainwright
Departments of Statistics, and
Electrical Engineering and Computer Science
University of California, Berkeley
wainwrig@eecs.berkeley.edu

Michael I. Jordan
Division of Computer Science, and
Department of Statistics
University of California, Berkeley
jordan@cs.berkeley.edu

September 15, 2004

Technical Report 671
Department of Statistics
University of California, Berkeley

Abstract

The Sherali-Adams (SA) and Lasserre (LS) approaches are "lift-and-project" methods that generate nested sequences of linear and/or semidefinite relaxations of an arbitrary 0-1 polytope $P \subseteq [0,1]^n$. Although both procedures are known to terminate with an exact description of P after n steps, there are various open questions associated with characterizing, for particular problem classes, whether exactness is obtained at some step $s < n$. This paper provides sufficient conditions for exactness of these relaxations based on the hypergraph-theoretic notion of treewidth. More specifically, we relate the combinatorial structure of a given polynomial system to an underlying hypergraph. We prove that the complexity of assessing the global validity of moment sequences, and hence the tightness of the SA and LS relaxations, is determined by the *treewidth* of this hypergraph. We provide some examples to illustrate this characterization.

Keywords: Integer programming; polynomial programs; semidefinite relaxation; linear programming; moment polytopes; graphical models; treewidth.

1 Introduction

Given a discrete set $F \subseteq \{0,1\}^n$, it is frequently of interest to characterize the 0-1 polytope P given by the convex hull of F in terms of a set of linear inequality constraints. In particular, such a characterization leads to a linear programming formulation of any optimization problem over F with a linear cost function. The field of polyhedral combinatorics [15] is concerned with finding efficient representations of such polytopes for various combinatorial problems.

Given some relaxation of the convex hull $\operatorname{conv} F$, there exist a variety of classical techniques for strengthening the relaxation via the addition of so-called "cutting planes" (e.g., Gomory-Chvátal cuts [3]). A second class of methods is based on the observation that it is often possible to represent the polytope P as the projection of another polytope Q lying in a higher-dimensional space. The basic idea is that projection can lead to a substantial increase in the number of facets; indeed, it can be possible to characterize a polytope P with exponentially many facets as the projection of a polytope Q with only a polynomial number of facets. Of course, such representations are of most interest when the dimension of Q remains polynomial in the dimension of P, which is known as a *compact representation* of P.

This possibility—namely, of representing P in terms of a projection—motivates the class of "lift-and-project" methods, in which constraints are imposed by first lifting the problem to higher dimension and then projecting back down. The past fifteen years have witnessed the development of a number of such techniques for constructing projection-based representations of general 0-1 polytopes,[1] including the Sherali-Adams procedure [16], the Lovász-Schrijver procedure [13], and most recently the Lasserre procedure [8]. Typically, these methods are defined in terms of a sequence of relaxations, indexed by the number of lifting steps applied. When applied to an 0-1 polytope P, all three methods are known to terminate with an exact description of P after at most n steps. This n^{th} order representation of P entails lifting to a space of dimension $\mathcal{O}(2^n)$, and hence is primarily of theoretical interest. Although the full n steps are required in the worst case, a variety of open questions concern whether fewer steps of lifting may suffice for problems with special structure.

The minimal number of lifting steps required is known as the *rank* of the relaxation. The rank issue has been studied for various specific problems in combinatorial optimization, including the matching problem [17], the stable set problem [e.g., 13, 11], and the MAX-CUT problem [e.g., 5, 10]. All of these problems can be viewed as particular types of *0-1 polynomial programs*, as defined in Section 2, with special structure in both the constraint set and the form of the objective function.

In this paper, we study the rank of the Sherali-Adams (henceforth SA) and Lasserre (henceforth LS) sequence of relaxations for general 0-1 polynomial programs. Our primary contribution is to define the *canonical hypergraph* associated with a polynomial program, and to prove that the rank of the SA and LS relaxations is upper bounded by the *treewidth* of this hypergraph. The cornerstone of our development is the *junction tree theorem* [4, 12], which provides a general framework for representing and manipulating probability distributions that factorize over hypergraphs. The junction tree formalism also leads to dynamic-programming algorithms defined on hypergraphs, which generalize and unify a wide class of standard algorithms for linearly ordered problems (e.g., Kalman filtering, Viterbi algorithm). A key consequence of the junction tree theorem—and one which we exploit in this paper—is a set of necessary and sufficient conditions for the global consistency of moment sequences. We show that these conditions, in the context of the canonical hypergraph defined by a polynomial system, can be used to assess the exactness of the SA and LS relaxations. Moreover, the resulting characterization of exactness is sharp in a worst-case sense that we make precise. Finally, we discuss some examples to illustrate this treewidth-based characterization.

2 Hierarchies of semidefinite relaxations

In this section, we begin with the basic set-up necessary to describe the Sherali-Adams (SA) and Lasserre (LS) procedures. In doing so, we follow largely the notation and approach of Lasserre [8] and Laurent [11].

[1] In fact, the underlying ideas are applicable more broadly to general semi-algebraic sets, as described in the work of Lasserre [9] and Parrilo [14].

2.1 Notation and set-up

Let $\mathcal{V} = \{1, \ldots, n\}$, and let $\mathcal{P}(\mathcal{V})$ denote its *power set*—i.e., the set of all subsets of $\mathcal{V}$. For any positive integer $q \leq n$, let $\mathcal{P}_q(\mathcal{V})$ denote the set of all subsets of $\mathcal{V}$ of cardinality at most q. Of central importance are constraint sets described by polynomials in $\{x_1, \ldots, x_n\}$, in which each x_i has degree either zero or one. Any such polynomial g can be written in the form

$$g(x) \quad := \quad \sum_{I \subseteq \mathcal{V}} g[I] x[I], \tag{1}$$

where $x[I] := \prod_{i \in I} x_i$, and $\{g[I],\ I \subseteq \mathcal{V}\}$ are the coefficients associated with this monomial expansion. Our convention is to set $x[\emptyset] = 1$.

Let $g_1, \ldots, g_m$ be a collection of such polynomials, and consider the semi-algebraic set described by enforcing non-negativity of these polynomials over the hypercube:

$$K \quad := \quad \{x \in [0,1]^n \mid g_\ell(x) \geq 0 \ \text{ for } \ \ell = 1, \ldots, m\}. \tag{2}$$

When necessary, we write $K(g_1, \ldots, g_m)$ to emphasize the fact that K is defined by the underlying polynomials. The following pieces of notation will be useful in the sequel: for each $\ell = 1, \ldots, m$, let $w_\ell = \deg(g_\ell)$ and $v_\ell := \lceil w_\ell / 2 \rceil$. We also define $w := \max_{\ell=1,\ldots,m} w_\ell$ and $v := \max_{\ell=1,\ldots,m} v_\ell$.

Letting g_0 be another polynomial of the form (1), we consider 0-1 polynomial programs of the following type:

$$\min_{x \in \mathbb{R}^n} g_0(x) \quad \text{s.t.} \quad x \in \{0,1\}^n \cap K(g_1, \ldots, g_m). \tag{3}$$

It is often convenient to assume a linear cost function—say of the form $g_0(x) = \sum_{i=1}^n g_0[i]\, x_i$ for some real-valued coefficients $g_0[i]$. Note that this assumption entails no loss of generality: if the cost function involves a higher monomial—say $\prod_{i \in I} x_i$—it can be eliminated by introducing a new variable x_{n+1}, and then adding an extra equality constraint (i.e., a pair of inequalities) to enforce the relation $x_{n+1} - \prod_{i \in I} x_i = 0$. Although the bulk of our development assumes linear cost, we find it convenient to retain non-linear cost in certain cases (see, e.g., Section 5.1).

When the cost function is linear, problem (3) can be reformulated as the linear program

$$\min_{\mu \in \mathbb{R}^n} \sum_{i=1}^n g_0[i]\, \mu[i] \quad \text{s.t.} \quad \mu \in P(g_1, \ldots, g_m), \tag{4}$$

in which the constraint set $P(g_1, \ldots, g_m) := \operatorname{conv}\big(\{0,1\}^n \cap K(g_1, \ldots, g_m)\big)$ is given by the convex hull of all feasible solutions in the original problem (3). This equivalence demonstrates that the complexity of solving problem (3) is closely related to the structure of the polytope $P \equiv P(g_1, \ldots, g_m)$.

2.2 Moments and lift-and-project methods

A useful interpretation of the set P is as a *first-order moment polytope.* More specifically, the set P corresponds, by definition, to the set of all first-order moment vectors $\{\mu[i] = \mathbb{E}_p[x_i],\ i = 1, \ldots n\}$ that can be realized by a distribution p with support restricted to $\{0,1\}^n \cap K$. The key idea underlying lift-and-project methods is that any such first-order moment polytope can be characterized as the projection of a semidefinite constraint set that is specified in terms of *higher-order* moments.

Given a distribution p, let us define, for each subset $I \subseteq \mathcal{V}$, the associated multinomial moment

$$\mu[I] \;:=\; \mathbb{E}_p\{x[I]\} \;=\; \mathbb{E}_p\left\{\prod_{i \in I} x_i\right\}. \tag{5}$$

These multinomial moments form a real-valued vector $\mu = \{\mu[I],\ I \subseteq \mathcal{V}\}$ indexed by the power set $\mathcal{P}(\mathcal{V})$.

Given an arbitrary vector $y \in \mathbb{R}^{\mathcal{P}(\mathcal{V})}$, we now describe semidefinite constraints that can be used to assess whether or not it is a valid moment sequence for a distribution p with support constrained to $\{0,1\}^n \cap K$. We begin by using y to define a $2^n \times 2^n$ matrix $M_n[y]$, with rows and columns indexed by elements of $\mathcal{P}(\mathcal{V})$, as follows:

$$\big(M_n[y]\big)_{IJ} \;:=\; y[I \cup J]. \tag{6}$$

The structure of $M_n[y]$ for $n = 3$ is illustrated in Figure 1(a).

For each $\ell = 1, \ldots, m$, the polynomial $g_\ell(\cdot)$ is uniquely identified by its coefficient vector $g_\ell := \{g_\ell[I],\ I \subseteq \mathcal{V}\}$. With this notation, we define for each $\ell = 1, \ldots, m$ another $2^n \times 2^n$ matrix $M_n[g_\ell * y]$ as follows:

$$\big(M_n[g_\ell * y]\big)_{IJ} \;:=\; \sum_{U \subseteq \mathcal{V}} g_\ell[U]\, y[I \cup J \cup U]. \tag{7}$$

By imposing positive semidefiniteness constraints on these matrices, we obtain the following constraint set:

$$C_n(K) \;:=\; \{y \in \mathbb{R}^{\mathcal{P}(\mathcal{V})} \mid M_n[y] \succeq 0 \text{ and } M_n[g_\ell * y] \succeq 0 \text{ for all } \ell = 1, \ldots m\}. \tag{8}$$

For any set $A \subseteq \mathbb{R}^{\mathcal{P}(\mathcal{V})}$, let $\Pi_1(A)$ denote its projection onto the n co-ordinates indexed by the sets of cardinality one. In terms of moments, this operation amounts to projecting onto the n-dimensional vector $\mu[1], \ldots, \mu[n]$ of first-order moments. With this notation, it was proved first by Lasserre [8], and subsequently in an alternative way by Laurent [11], that the following equivalence holds:

$$P(g_1, \ldots, g_m) \;=\; \Pi_1\Big[C_n(K) \;\cap\; \big\{y \in \mathbb{R}^{\mathcal{P}(\mathcal{V})} \mid y[\emptyset] = 1\big\}\Big]. \tag{9}$$

The essence of this result is that the first-order moment polytope P can be characterized by a "lift-and-project" procedure, which entails lifting the problem by introducing higher-order moments, then imposing constraints in the lifted space, and finally projecting back to the co-ordinates corresponding to first-order moments.

Of course, the characterization of P given in equation (9) is not a practical one, since it requires imposing positive semidefiniteness constraints on matrices of size $2^n \times 2^n$. However, this equivalence does suggest that outer bounds on P can be obtained by imposing positive semidefiniteness on particular minors of the matrices $M_n(y)$ and $M_n(g_\ell * y)$. Moreover, as shown by Laurent [11], this perspective leads to a unified understanding of both the Lasserre and Sherali-Adams hierarchies.

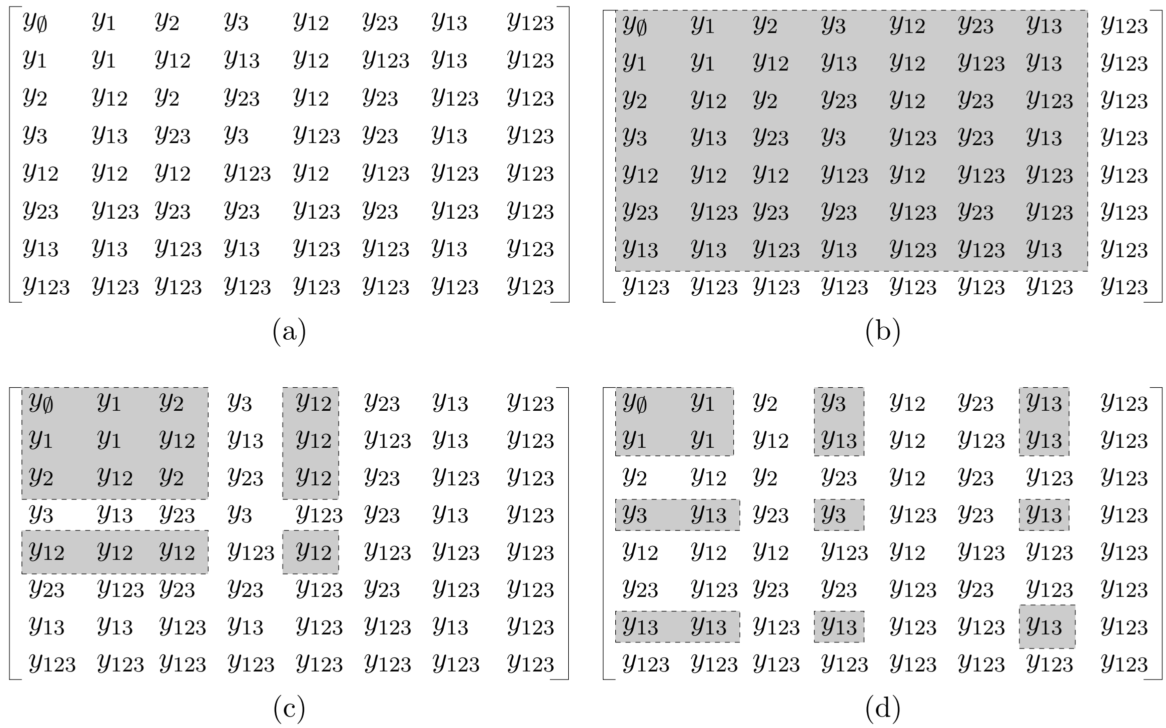

Figure 1. Relevant minors for the Lasserre and Sherali-Adams relaxations. (a) Full matrix $M_3[y]$. (b) Shaded region: 7×7 principal minor $M_2[y]$ constrained by the Lasserre relaxation at order 1. (c), (d) Shaded regions: 4×4 minors $M_{\{12\}}[y]$ and $M_{\{13\}}[y]$ constrained by the Sherali-Adams relaxation at order 2. Also constrained is the minor $M_{\{23\}}[y]$ (not shown).

2.3 Lasserre hierarchy

The Lasserre hierarchy of relaxations entails imposing positive semidefiniteness on certain principal minors of the matrices $M_n[y]$ and $M_n[g_\ell * y]$. For any integer $1 \leq s \leq n$, let $M_s[y]$ denote the principal minor of $M_n[y]$ indexed by $\mathcal{P}_s(\mathcal{V})$. In explicit terms, this matrix $M_s[y]$ has elements of the form

$$M_s[y] \quad := \quad \big(y[I \cup J]\big)_{|I|,|J| \leq s}.$$

For $n = 3$, the matrix $M_2[y]$ is shown in Figure 1(b).

Recall the definitions $v_\ell := \lceil \deg(g_\ell)/2 \rceil$ and $v := \max\limits_{\ell=1,\ldots,m} v_\ell$. For each integer $q \geq v-1$, define the set

$$\widetilde{L}_q(K) := \big\{ y \in \mathbb{R}^{\mathcal{P}_{2q+2}(\mathcal{V})} \;\big|\; M_{q+1}[y] \succeq 0 \text{ and } M_{q+1-v_l}[g_\ell * y] \succeq 0 \text{ for all } \ell = 1, \ldots m \big\}. \tag{10}$$

The *Lasserre relaxation* of order q is defined by intersecting the first co-ordinate with the hyperplane $y[\emptyset] = 1$, and then projecting the resulting set onto the co-ordinates associated with first-order moments:

$$L_q(K) \quad := \quad \Pi_1 \Big[\widetilde{L}_q(g_1, \ldots, g_m) \;\cap\; \big\{ y \in \mathbb{R}^{\mathcal{P}_{2q+2}(\mathcal{V})} \;\big|\; y[\emptyset] = 1 \big\} \Big]. \tag{11}$$

This construction generates a nested sequence of the form

$$P \;=\; L_{n+v-1}(K) \;\subseteq\; \ldots \;\subseteq\; L_v(K) \;\subseteq\; L_{v-1}(K). \tag{12}$$

2.4 Sherali-Adams hierarchy

The Sherali-Adams hierarchy [16] can also be understood in terms of imposing positive semidefiniteness on certain minors of $M_n[y]$ and $M_n[g_\ell * y]$. For any subset $U \subseteq \mathcal{V}$, we isolate the minor of $M_n[y]$ indexed by subsets of U as follows:

$$M_U[y] \;\; := \;\; \big(y[I \cup J]\big)_{I,J \subseteq U}$$

Note that, with this notation, we have the equivalence $M_{\mathcal{V}}[y] = M_n[y]$. For $n = 3$, the matrices $M_{\{12\}}[y]$ and $M_{\{13\}}[y]$ are shown in panels (c) and (d) respectively of Figure 1.

Recall the definition $w := \max_{\ell=1,\ldots,m} \deg(g_\ell)$. For each $q = 1, \ldots, n$, we define the set $\widetilde{S}_q(K)$ as follows:

$$\left\{ y \in \mathbb{R}^{\mathcal{P}_{q+w}(\mathcal{V})} \;\middle|\; \begin{array}{c} M_W[y] \succeq 0, \text{ for all } W \subseteq \mathcal{V} \text{ such that } |W| = \min(q+1, n) \\ M_U[g_\ell * y] \succeq 0, \text{ for all } U \subseteq \mathcal{V} \text{ such that } |U| = q, \;\; \forall \ell = 1, \ldots m \end{array} \right\}. \tag{13}$$

In analogy to the Lasserre construction, the *Sherali-Adams* relaxation of order q is defined by intersecting the first co-ordinate with the hyperplane $y[\emptyset] = 1$, and then projecting the resulting set onto the co-ordinates associated with first-order moments:

$$S_q(K) \;\; := \;\; \Pi_1\Big[\widetilde{S}_q(K) \;\cap\; \big\{y \in \mathbb{R}^{\mathcal{P}_{q+w}(\mathcal{V})} \;\big|\; y[\emptyset] = 1\big\}\Big], \tag{14}$$

thereby generating the nested sequence of relaxations:

$$P \;=\; S_n(K) \;\subseteq\; \ldots \;\subseteq\; S_2(K) \;\subseteq\; S_1(K). \tag{15}$$

3 Hypertrees and the junction tree representation

In this section, we introduce background material on hypergraphs, hypertrees and the junction tree representation that underlies our development in the sequel. Further background on hypergraphs, junction trees and treewidth can be found in various sources [1, 2, 4, 12].

3.1 Hypergraphs and hypertrees

A *hypergraph* $\mathcal{H} = (\mathcal{V}, \mathcal{E})$ consists of a set $\mathcal{V}$ of *vertices* and a set $\mathcal{E}$ of *hyperedges*. Each hyperedge corresponds to a particular subset of $\mathcal{V}$ (i.e., an element of the power set $\mathcal{P}(\mathcal{V})$). For the sake of exposition, it is convenient to adopt the convention that for any hyperedge $E \in \mathcal{E}$, the hyperedge set also contains all $F \subseteq E$. We say that a hyperedge is *maximal* if it is not properly contained within any other hyperedge. Under our convention, we necessarily have $\mathcal{E} = \cup_\alpha \mathcal{P}(E_\alpha)$, where the union is taken over all maximal hyperedges. Finally, note that our definition of a hypergraph covers as a special case an ordinary graph, for which all of the maximal hyperedges have cardinality two.

A fundamental class of graphs are those without cycles, known as trees or acyclic graphs. The following definition extends this notion to hypergraphs.

Definition 1 (Tree decomposition). A *tree decomposition* of a hypergraph $\mathcal{H} = (\mathcal{V}, \mathcal{E})$ is a tree-structured graph $\mathcal{T}$ with the following properties:

(a) the vertex set of the tree $\mathcal{T}$ is of the form $\mathcal{V}_{\mathcal{T}} = \{E_1, \ldots, E_d\}$, where each $E_\alpha \in \mathcal{E}$;

(b) every hyperedge $F \in \mathcal{E}$ is contained within some $E_\alpha \in \mathcal{V}_{\mathcal{T}}$, and

(c) for every pair of vertices $E_\alpha, E_\beta \in \mathcal{V}_{\mathcal{T}}$, the intersection $E_\alpha \cap E_\beta$ is contained in every $E_\gamma \in \mathcal{V}_{\mathcal{T}}$ on the unique path in $\mathcal{T}$ that joins E_α and E_β. This property is known as *running intersection.*

The quantity $t := \max_{\alpha=1,\ldots d} |E_\alpha| - 1$ is known as the *width* of the tree decomposition. We refer to any hypergraph with a tree decomposition of width t as a *hypertree of width* t.

Example 1. (a) The simplest illustration is provided by an ordinary tree, for which the edges play the role of vertices in the tree decomposition. Since the size of each edge is two, an ordinary tree corresponds to a hypertree of width 1.

(b) The hypergraph on $\mathcal{V} = \{1, 2, 3, 4, 5\}$ with maximal hyperedges (124) and (235) is is a hypertree of width two.

While not every hypergraph is a hypertree, a suitable augmentation of the hyperedge set allows every hypergraph to be covered by a hypertree. More formally, we say that $\mathcal{G}' = (\mathcal{V}, \mathcal{E}')$ is a *covering hypergraph* for $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ if $\mathcal{E} \subseteq \mathcal{E}'$. Of interest will be the set of covering hypertrees associated with a given hypergraph $\mathcal{G}$. Clearly, every hypergraph has at least one covering hypergraph: more specifically, if we include $(12 \cdots n)$ as a hyperedge, then the trivial tree with vertex set $\mathcal{V}_{\mathcal{T}} = (12 \cdots n)$ is a covering hypertree of width $n - 1$ for any hypergraph. The key quantity, then, is the minimal width taken over all possible covering hypertrees, as we formalize in the following:

Definition 2 (Treewidth). The *treewidth* of a hypergraph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is the minimum width of all hypertrees $\mathcal{G}' = (\mathcal{V}, \mathcal{E}')$ that cover $\mathcal{G}$.

We illustrate this definition with an example:

Example 2. Consider the hypergraph with $\mathcal{V} = \{1, 2, 3, 4\}$ and $\mathcal{E} = \mathcal{V} \cup \{(12), (23), (34), (41)\})$, which corresponds to an ordinary single cycle graph on four vertices. It can be verified that no subset of $\mathcal{E}$ alone defines a tree decomposition. (For instance, using the subset $\{(12), (23), (34)\}$ as the tree vertex set $\mathcal{V}_{\mathcal{T}}$ satisfies properties (a) and (b) of the definition, but fails the running intersection property (c).) However, one minimal covering hypertree can be specified by maximal hyperedges $\{(123), (234)\}$; these hyperedges also define the vertex set $\mathcal{V}_{\mathcal{T}}$ of the associated tree decomposition. Accordingly, the treewidth of the single cycle graph is 2.

3.2 Junction tree representation

Let us say that a function $f : \{0, 1\}^{|S|} \to \mathbb{R}$ is *localized* to the subset $S \subset \mathcal{V}$ if f depends on only on elements within S. Now consider a probability distribution p over $\{0, 1\}^n$ that is composed as the product of non-negative functions that are localized to the hyperedges of a given hypergraph. The essence of the junction tree theorem [12] is that when the underlying hypergraph is a hypertree, then any such distribution has an alternative but equivalent factorization in terms of local marginal distributions defined on the maximal hyperedges (and their pairwise intersections). Moreover, this factorization can be computed efficiently by a variant of dynamic programming on the junction tree structure.

Central to our development in the sequel is a result that can be viewed as a corollary of the junction tree theorem [18]. Consider a hypertree $\mathcal{H} = (\mathcal{V}, \mathcal{E})$, and for each maximal hyperedge E, let p_E be a vector with $2^{|E|}$ elements. Of interest is when the collection of vectors $\{p_E \mid E \in \mathcal{E}\}$ correspond to a set of marginal distributions induced by some global distribution p on $\{0,1\}^n$. The junction tree theorem implies that for any hypertree $\mathcal{H}$, the following conditions are both necessary and sufficient to ensure that $\{p_E \mid E \in \mathcal{E}\}$ are the marginals of some global distribution:

1. Each vector p_E is non-negative.
2. Each vector p_E is properly normalized (i.e., $\|p_E\|_1 = 1$).
3. For any pair $E, F \in \mathcal{E}$ with non-empty intersection $E \cap F$, the marginals induced by p_E and p_F on variables in their intersection agree with each other.

This characterization of globally consistent marginal distributions plays a key role in the proof of our main result.

4 Exactness based on treewidth

We now proceed to the statement and proof of the main result.

4.1 Hypergraph of a polynomial system

Given a polynomial g of the form (1), we use $\mathcal{V}[g]$ to denote the subset of variables x_i on which g depends, defined as follows:

$$\mathcal{V}[g] \quad := \quad \big\{i \in \mathcal{V} \mid i \in I \subseteq \mathcal{V} \text{ with } g[I] \neq 0\big\}. \tag{16}$$

The *canonical hypergraph* associated with the polynomial system $(g_1, \ldots, g_m)$ consists of the vertex set $\mathcal{V} = \{1, \ldots, n\}$, and the hyperedge set

$$\mathcal{E} \quad := \quad \Big\{\mathcal{P}(\mathcal{V}[g_\ell]) \;\Big|\; \ell = 1, \ldots, m\Big\}. \tag{17}$$

With these definitions, we have the following:

Theorem 1. *Let t be the treewidth of the canonical hypergraph associated with the polynomial system $(g_1, \ldots, g_m)$. Then:*

(a) The Sherali-Adams relaxation is upper bounded by $t+1$ (i.e., $S_{t+1}(K) = P$).

(b) The Lasserre relaxation is upper bounded by $t+v$ (i.e., $L_{t+v}(K) = P$).

Moreover, these bounds are sharp in the following worst-case sense: for a given treewidth t, there exists a polynomial system $(g_1, \ldots, g_m)$ whose associated canonical hypergraph has treewidth t such that the upper bounds in (a) and (b) are both met with equality.

Remarks:

(a) To gain some intuition for the theorem, consider a simple example with $\mathcal{V} = \{1, 2, 3, 4, 5\}$. Suppose that we impose the polynomial constraints $g_1(x_1, x_2, x_4) = x_1 x_2 x_4 - 1 \geq 0$ and $g_2(x_2, x_3, x_5) = x_2 x_3 x_5 = 0$. The associated canonical hypergraph has maximal hyperedges $\{1, 2, 4\}$ and $\{2, 3, 5\}$, which generates the hypertree of width 2 considered in Example 1(a). Therefore, by Theorem 1(a), the Sherali-Adams procedure will be exact at order 3.

(b) It should be noted that the conditions of Theorem 1 are sufficient to ensure exactness, but not all of the associated constraints may be necessary in general. This comment is particularly applicable to the Lasserre relaxation. For instance, it is known [11] that the inclusion $L_{t+v}(K) \subseteq S_{t+1}(K)$ always holds. In general, however, the Sherali-Adams relaxation $S_{t+1}(K)$ is based on fewer constraints than the Lasserre relaxation $L_{t+v}(K)$. Therefore, when equality holds in both relaxations, there must be irrelevant constraints involved in the Lasserre construction. We elaborate on the issue of the minimal number of constraints in Section 4.3.

4.2 Proof of Theorem 1

We begin by introducing a definition and lemma that are central to the proof:

Definition 3. Let $\mathcal{S}$ be a subset of $\{0, 1\}^n$. Given some subset $\mathcal{Q}$ of the power set $\mathcal{P}(\mathcal{V})$, let $\mathbf{y}\{\mathcal{Q}\}$ denote a vector indexed by the elements of $\mathcal{Q}$ (i.e., of the form $\{y[I] \mid I \in \mathcal{Q}\}$). We say that $\mathbf{y}\{\mathcal{Q}\}$ is a *globally consistent $\mathcal{S}$-moment vector for $\mathcal{Q}$* if there exists some distribution p with support restricted to $\mathcal{S} \subseteq \{0, 1\}^n$ such that $y[I] = \mathbb{E}_p\{x[I]\}$ for all $I \in \mathcal{Q}$.

The following lemma is proved in Appendix A:

Lemma 1 (Local to global consistency). *Consider a hypertree $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ of width t. For each maximal hyperedge $E \in \mathcal{E}$, let $\mathcal{S}(E)$ be some subset of $\{0, 1\}^{|E|}$, and consider the subset $\mathcal{S} \subseteq \{0, 1\}^n$ defined by*

$$\mathcal{S} \quad := \quad \cap_{E \in \mathcal{E}} \big\{ u \in \{0, 1\}^n \mid u_E \in \mathcal{S}(E) \big\}. \tag{18}$$

where $u_E := \{u_i \mid i \in E\}$. Then a vector $\mathbf{y}\{\mathcal{E}\}$ is a globally consistent $\mathcal{S}$-moment vector for $\mathcal{E}$ if and only if the subvector $\mathbf{y}\{\mathcal{P}(E)\}$ is a globally consistent $\mathcal{S}(E)$-moment vector for each maximal hyperedge $E \in \mathcal{E}$.

Equipped with this lemma, we first prove part (a) of the theorem. Consider the canonical hypergraph associated with the polynomial system $(g_1, \ldots, g_m)$. Since it has width t by assumption, there exists a covering hypertree $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with width t. Let $\mathcal{V}_T = \{E_1, \ldots, E_d\}$ be the vertex set of an associated tree decomposition, for which (by definition of treewidth) there holds $\max_{\alpha=1,\ldots,d} |E_\alpha| = t+1$. Since the hypertree covers the canonical hypergraph, for each $\ell = 1, \ldots, m$, the set $\mathcal{V}[g_\ell]$ is contained in at least one $E_\alpha \in \mathcal{V}_T$. For each $\alpha = 1, \ldots, d$, we define

$$G(E_\alpha) := \Big\{ \ell \in \{1, \ldots, m\} \mid \mathcal{V}[g_\ell] \subseteq E_\alpha \Big\}.$$

Note that by definition, for each $\ell \in G(E_\alpha)$, the value of the polynomial g_ℓ depends only on (at most) the subvector $u_{E_\alpha} := \{u_i \mid i \in E_\alpha\}$. With these definitions, the set $K \cap \{0,1\}^n$ has the decomposition

$$\cap_{\alpha=1}^{d} \big\{ u \in \{0,1\}^n \mid g_\ell(u_{E_\alpha}) \geq 0 \;\; \forall \, \ell \in G(E_\alpha) \big\} \;\; \equiv \;\; \cap_{\alpha=1}^{d} \mathcal{S}(E_\alpha), \tag{19}$$

where we have defined for future reference the local support set

$$\mathcal{S}(E_\alpha) \;\; := \;\; \big\{ u_{E_\alpha} \in \{0,1\}^{|E_\alpha|} \; \big| \; g_\ell(u_{E_\alpha}) \geq 0 \;\; \forall \, \ell \in G(E_\alpha) \big\}. \tag{20}$$

By definition, the polytope P can be characterized exactly by a suitable projection of the set of all vectors $\mathbf{y}\{\mathcal{E}\}$ that are globally valid $K \cap \{0,1\}^n$-moment vectors. Consider the Sherali-Adams relaxation $\widetilde{S}_{t+1}(K)$, as defined in equation (13). Since any hyperedge $E_\alpha \in \mathcal{E}$ has size at most $t+1$, any $\mathbf{y} \in \widetilde{S}_{t+1}(K)$ can be used to define a subvector $\mathbf{y}\{\mathcal{E}\}$. To establish exactness of the Sherali-Adams relaxation at order $t+1$, it suffices to show that any subvector $\mathbf{y}\{\mathcal{E}\}$ thus defined is a globally valid $K \cap \{0,1\}^n$-moment vector. Recall that $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is a hypertree, and the decomposition of $K \in \{0,1\}^n$ in equation (19). These facts in conjunction with Lemma 1 guarantee that $\mathbf{y}\{\mathcal{E}\}$ is a globally valid $K \cap \{0,1\}^n$-moment vector for $\mathcal{E}$ if and only if each of the subvectors $\mathbf{y}\{\mathcal{P}(E_\alpha)\}$, for $\alpha = 1, \ldots, d$, is a globally consistent $\mathcal{S}(E_\alpha)$-moment vector for $\mathcal{P}(E_\alpha)$.

Fix an arbitrary index $\alpha \in \{1, \ldots, d\}$. Since E_α has cardinality at most $t+1$, characterizing the set of all globally consistent $\mathcal{S}(E_\alpha)$-moment vectors is a subproblem for which the Sherali-Adams relaxation, if applied to E_α as an isolated subsystem, will be tight at order $t+1$. We now claim that the Sherali-Adams relaxation applied globally to the full set $\mathcal{V}$ at order $t+1$ imposes all of the necessary constraints to ensure exactness for the E_α subsystem. More precisely, since each E_α has cardinality at most $t+1$, the global Sherali-Adams relaxation at order $t+1$ imposes the constraint $M_{E_\alpha}[\mathbf{y}\{\mathcal{P}(E_\alpha)\}] \succeq 0$, as well as the constraints $M_{E_\alpha}[g_\ell * y] \succeq 0$ for all $\ell \in G(E_\alpha)$. These conditions guarantee that $\mathbf{y}\{\mathcal{P}(E_\alpha)\}$ is a globally consistent $\mathcal{S}(E_\alpha)$-moment vector for $\mathcal{P}(E_\alpha)$, which concludes the proof of part (a).

To establish part (b), we use the fact that $L_{q+v-1}(K) \subseteq S_q(K)$ for any $q = 1, \ldots, n$, as established by Laurent [11]. Part (b) thus follows from part (a) with $q = t+1 \leq n$.

Finally, the worst-case tightness of the bounds follows as a consequence of the junction tree theorem. If we are allowed freedom in our choice of polynomial system $(g_1, \ldots, g_m)$ of treewidth t, then the full set of junction tree constraints are required for an exact characterization. The preceding proof establishes the number of steps required to impose all of the junction tree constraints. □

4.3 Minimal set of constraints

The proof of Theorem 1 actually provides somewhat more precise information on the minimal set of constraints required to ensure global validity of a moment sequence, and hence tightness of the associated relaxation. In general, both the SA and LS sequences impose more constraints than are strictly necessary for a given treewidth. In order to illustrate, suppose that the polynomials defining K are all of degree one (so that $v = 1$), and the canonical hypergraph corresponding to a given polynomial system is a hypertree of width one (i.e., an ordinary tree). As a particular example, consider $n = 3$ and the constraints $g_1(x_1, x_2) = x_1 + x_2 - 1 \geq 0$ and $g_2(x_2, x_3) = 1 - x_2 - x_3 \geq 0$. The associated canonical hypergraph is the ordinary tree with edge set $\{(12), (23)\}$.

In this tree-structured case, Theorem 1 asserts that both the SA and LS sequences are tight at order 2, which is certainly true. By definition, the SA sequence at order 2 will enforce, among other constraints, positive semidefiniteness (PSD) on all submatrices $M_U[y]$ with $|U| = 3$ and all submatrices $M_W[g_\ell * y]$ with $|W| = 2$. However, for a hypertree of width one, careful inspection of the proof of Theorem 1 reveals that it suffices to impose PSD constraints only for subsets U or W with cardinality two, and moreover only for those subsets corresponding to the hyperedges. To illustrate with our specific example, it suffices to impose the constraints $M_U[y] \succeq 0$ and $M_W[y] \succeq 0$ for U and W ranging only over the edge set $\{(12), (13)\}$. Similar comments apply to the Lasserre sequence. Therefore, Theorem 1 shows that for a hypergraph of given treewidth, a subset of the constraints in both the SA and LS sequences are redundant.

5 Illustrative examples

We illustrate Theorem 1 with some examples.

5.1 0-1 quadratic programs

Suppose that we are interested in an arbitrary 0-1 quadratic program (QP) of the form

$$\max_{x \in \{0,1\}^m} \Big\{ \sum_{i=1}^m q_i x_i + \sum_{i<j} q_{ij} x_i x_j \Big\}. \tag{21}$$

One approach is to reduce this QP to the canonical form (3) with a linear cost function by defining a set of $\binom{n}{2}$ additional variables $\{x_{ij} \mid 1 \le i < j \le n\}$, and imposing equality constraints of the form $x_{ij} - x_i x_j = 0$.

If we follow this route, the canonical hypergraph associated with the quadratic program will have hyperedges of size three; in particular, the set of maximal hyperedges will be

$$\Big\{ \{i, j, (ij)\} \mid \text{ for all } (i,j) \text{ such that } q_{ij} \neq 0 \Big\}.$$

An alternative approach, and one which turns out to be more natural, is to retain the monomials $x_i x_j$ in their original form, and consider the problem of characterizing the convex hull of the vectors $\{(x_i)_{i=1}^n,\ (x_i x_j)_{i<j} \mid (x_i)_{i=1}^n \in \{0,1\}^n\}$. In terms of moments, the associated problem corresponds to characterizing not only the n-vector of singleton moments $y[i] = \mathbb{E}[x_i]$ for $i = 1, \ldots, n$, but also the $\binom{n}{2}$ pairwise moments $y[(ij)] = \mathbb{E}[x_i x_j]$ for $i < j$. It is straightforward to modify the proof of Theorem 1(a) to show that the Sherali-Adams relaxation will be exact at order $t+1$, where t is the treewidth of the ordinary graph $\mathcal{G}$ with vertex set $\mathcal{V} = \{1, \ldots, n\}$ and edge set $\mathcal{E} = \{\ (i,j) \mid q_{ij} \neq 0\}$.

It is worth commenting that this treewidth-based characterization takes into account only the graphical structure associated with the QP (21), but not the numerical values of the cost coefficients q_i and q_{ij}. With particular restrictions on the cost coefficients, the treewidth condition can be relatively weak. As an example, suppose that $q_i = 0$ for all $i = 1, \ldots, n$ and that $q_{ij} \neq 0$ only for edges (i,j) associated with a planar graph G. Up to some irrelevant constant terms, this problem is equivalent to a MAX-CUT problem on the associated graph. For a planar graph, it is known that the Sherali-Adams relaxation at order 3 (which imposes the triangle constraints that define the so-called metric polytope [5]) provides an exact description of the MAX-CUT problem. In contrast, the treewidth of a planar graph need not be bounded in n; for example, the four-nearest-neighbor grid in two dimensions with n nodes has a treewidth that scales as $\sqrt{n}$.

5.2 Rank n example

We now consider a problem taken from Laurent [11] for which n iterations of the Sherali-Adams procedure are required to characterize P exactly. Letting $\mathcal{V} = \{1, \ldots, n\}$, we define

$$K := \{x \in [0,1]^n \mid \sum_{r \in R}(1 - x_r) + \sum_{r \in \mathcal{V} \setminus R} x_r \geq \frac{1}{2} \quad \text{for all} \quad R \in \mathcal{P}(\mathcal{V})\}. \tag{22}$$

On one hand, it is straightforward to see that $K \cap \{0,1\}^n = \emptyset$, and hence $P := \operatorname{conv}\left[K \cap \{0,1\}^n\right]$ is also empty. On the other hand, Laurent [11] shows that the set $S_{n-1}(K)$ is not empty, implying that the Sherali-Adams procedure requires the maximal n iterations.

With reference to Theorem 1, the canonical hypergraph associated with problem (22) consists of the single maximal hyperedge $E_1 = \{1, \ldots, n\}$, and hence has treewidth $n - 1$. Thus, the fact that n iterations are required is consistent with the assertion of Theorem 1, and the upper bound on the rank is met with equality.

6 Conclusion

The main contribution of this paper is to provide upper bounds on the ranks of the Sherali-Adams [16] and Lasserre [8] sequences of relaxations that are sharp in a worst-case sense. The central underlying ideas are that of the canonical hypergraph associated with a polynomial program, and the link provided by the junction tree theorem [12] between treewidth and the global consistency of moment sequences. More broadly, this paper establishes a connection between methods for solving polynomial programs based on moment sequences (or from the dual perspective, sums-of-squares representation), and the theory of hypergraphs. Polynomial programs of interest in practical applications will often exhibit special graphical features such as sparsity (e.g., due to spatially localized interactions). Thus, the high-level perspective taken in this paper—namely of exploiting the graphical structure of given polynomial systems—has broader consequences for characterizing the behavior of moment-based relaxations in applications (e.g., error-control coding [7, 6]).

A Proof of Lemma 1

The necessity is clear, since each $\mathbf{y}\{\mathcal{P}(E)\}$ is a subvector of $\mathbf{y}\{\mathcal{E}\}$ by definition. To establish sufficiency, suppose that $\mathbf{y}\{\mathcal{P}(E)\}$ is a globally consistent $\mathcal{S}(E)$-moment vector for each maximal $E \in \mathcal{E}$. To simplify matters, we restrict our attention to the hyperedges in the vertex set $\mathcal{V}_{\mathcal{T}} = \{E_1, \ldots, E_\alpha\}$ of some tree decomposition of the hypertree. By definition, for each $\alpha = 1, \ldots, d$, there exists a distribution p_α with support $\mathcal{S}(E_\alpha)$ such that $y[F] = \mathbb{E}_{p_\alpha}\big\{x[F]\big\}$ for each $F \subseteq E_\alpha$. In fact, the distribution p_α is unique since the inclusion-exclusion formula provides a one-to-one mapping between the $2^{|E_\alpha|}$-vector $\mathbf{y}\{\mathcal{P}(E_\alpha)\}$ and the $2^{|E_\alpha|}$-vector p_α. (See Laurent [11] for details of this mapping.) Now consider the distributions p_α and p_β for any pair (α, β) such that $E_\alpha \cap E_\beta \neq \emptyset$. We claim that the marginal distributions induced by p_α and p_β on their overlap are identical. Each of these induced marginal distributions has $2^{|E_\alpha \cap E_\beta|} - 1$ degrees of freedom (where we lose one degree due to the normalization constraint). Moreover, by construction of p_α and p_β, there holds $\mathbb{E}_{p_\alpha}\{x[F]\} = \mathbb{E}_{p_\beta}\{x[F]\}$ for all non-empty $F \subseteq E_\alpha \cap E_\beta$, which corresponds to a set of $2^{|E_\alpha \cap E_\beta|} - 1$ linearly independent constraints linking the marginals induced by p_α and p_β, so that they must agree.

Thus, for each $E_\alpha, \alpha = 1, \ldots, d$ in the tree decomposition, we have constructed a local distribution p_α over $\mathcal{S}(E_\alpha)$, such that p_α and p_β agree on their overlap $E_\alpha \cap E_\beta$. By the junction tree theorem, the local distributions $\{p_1, \ldots, p_d\}$ can be used to construct an unique global distribution p over $\{0,1\}^n$ such that $y[F] = \mathbb{E}_p\{x[F]\}$ for all $F \in \mathcal{E}$. This establishes that $\mathbf{y}\{\mathcal{E}\}$ is a globally consistent $\{0,1\}^n$-moment vector.

To conclude the proof, we need to verify that p has support $\mathcal{S}$. By the junction tree construction, for each $\alpha = 1, \ldots, d$, the distribution p, when marginalized down to the subset E_α, agrees with p_α. Suppose that p placed strictly positive mass on any configuration $u \notin \mathcal{S}$. By the definition of $\mathcal{S}$, at least of the marginals p_α would have mass outside $\mathcal{S}(E_\alpha)$, which is not possible by construction of p_α. Therefore, p must have support restricted to $\mathcal{S}$ so that $\mathbf{y}\{\mathcal{E}\}$ is a globally consistent $\mathcal{S}$-moment vector. □

References

[1] C. Berge. *Graphs and Hypergraphs.* North-Holland Publishing Company, Amsterdam, 1976.

[2] H. Bodlaender. A tourist guide through treewidth. *Acta Cybernetica*, 11:1–21, 1993.

[3] V. Chvátal. Edmonds polytopes and a hierarchy of combinatorial problems. *Discrete Mathematics*, 4:305–377, 1973.

[4] R. G. Cowell, A. P. Dawid, S. L. Lauritzen, and D. J. Spiegelhalter. *Probabilistic Networks and Expert Systems.* Statistics for Engineering and Information Science. Springer-Verlag, 1999.

[5] M. Deza and M. Laurent. *Geometry of Cuts and Metric Embeddings.* Springer-Verlag, New York, 1997.

[6] J. Feldman, T. Malkin, R. A. Servedio, C. Stein, and M. J. Wainwright. LP decoding corrects a constant fraction of errors. Technical Report CORC Technical Report TR-2003-08, Operations Research, Columbia University, December 2003.

[7] J. Feldman, M. J. Wainwright, and D. R. Karger. Using linear programming to decode binary linear codes. Accepted to IEEE Transactions on Information Theory, 2004.

[8] J. B. Lasserre. An explicit equivalent positive semidefinite program for nonlinear 0-1 programs. *SIAM Journal on Optimization*, 12:756–769, 2001.

[9] J. B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11(3):796–817, 2001.

[10] M. Laurent. Semidefinite relaxations for max-cut. In *The Sharpest Cut: Fetschrift in Honor of M. Padberg's 60th Birthday*, 2002.

[11] M. Laurent. A comparison of the Sherali-Adams, Lovász-Schrijver and Lasserre relaxations for 0-1 programming. *Mathematics of Operations Research*, 28:470–496, 2003.

[12] S. L. Lauritzen. *Graphical Models.* Oxford University Press, Oxford, 1996.

[13] L. Lovász and A. Schrijver. Cones of matrices and set-functions and $0-1$ optimization. *SIAM Journal of Optimization*, 1:166–190, 1991.

[14] P. Parrilo. Semidefinite programming relaxations for semialgebraic problems. *Mathematical Programming, Ser. B*, 96:293–320, 2003.

[15] A. Schrijver. *Combinatorial Optimization: Polyhedra and Efficiency.* Springer-Verlag, New York, NY, 2003.

[16] H. D. Sherali and W. P. Adams. A hierarchy of relaxations between the continuous and convex hull representations for zero-one programming problems. *SIAM Journal on Discrete Mathematics*, 3:411–430, 1990.

[17] T. Stephen and L. Tuncel. On a representation of the matching polytope via semidefinte liftings. *Mathematics of Operations Research*, 24:1–7, 1999.

[18] M. J. Wainwright and M. I. Jordan. Graphical models, exponential families, and variational inference. Technical report, UC Berkeley, Department of Statistics, No. 649, September 2003.